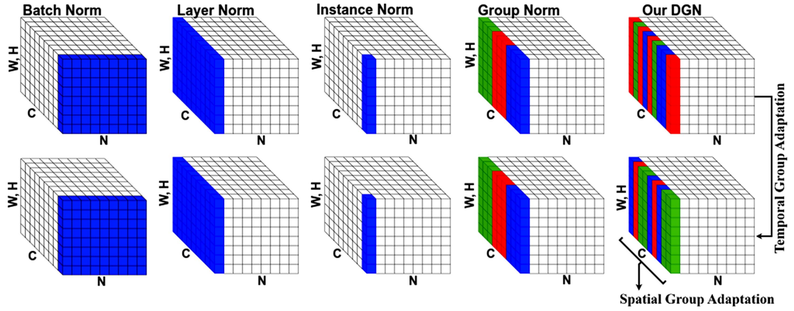
Dynamic Group Normalization: Spatio-Temporal Adaptation to Evolving Data Statistics

Yair Smadar and Assaf Hoogi

School of Computer Science, Ariel University

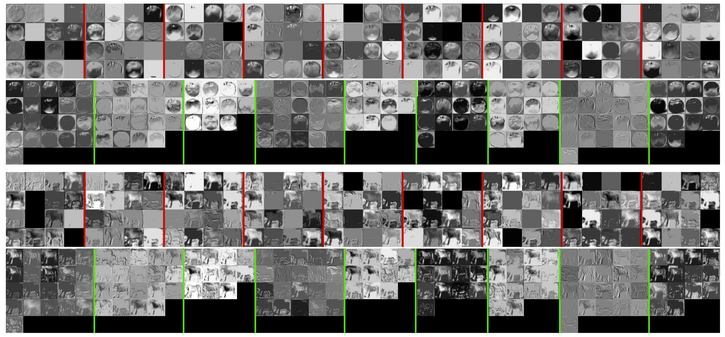
Motivation

- A fundamental limitation of current normalization techniques is their reliance on **rigid, fixed** normalization set sizes, which hinders adaptability to diverse or evolving statistical distributions.



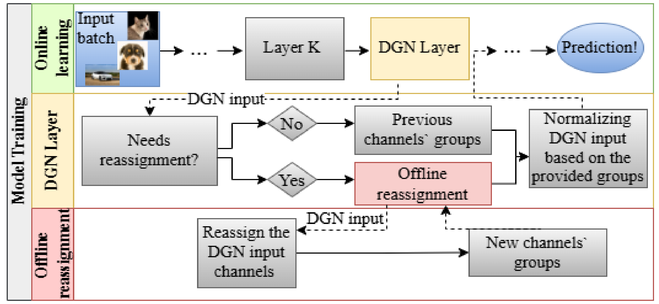
Our **Dynamic Group Normalization (DGN)** - **first adaptive framework** that

- **Dynamically** forms channel groups based on **statistical awareness.**
- Adapts across **spatial and temporal** dimensions.
- Integrates **offline** regrouping with **online learning.**



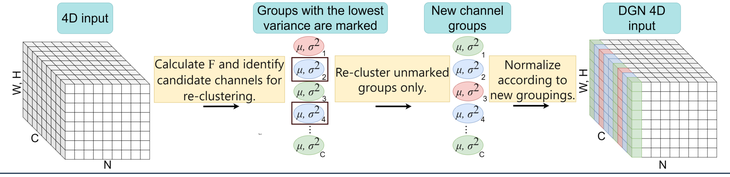
Red – vanilla GN, **Green** – our proposed DGN

Method



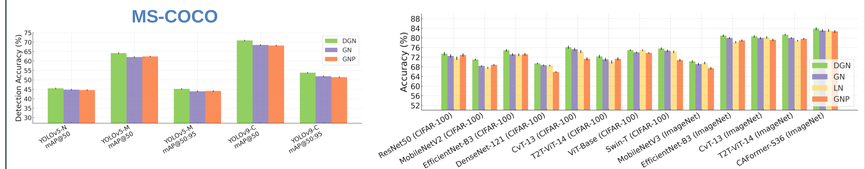
Channel Groups' Re-assignment

- Number of groups $\boldsymbol{K} = \left\lfloor \frac{C}{C_{G_{def}}} \right\rfloor$, $\boldsymbol{C}$ – total number of channels, $\boldsymbol{G_{def}}$ - initial default group size.
- Intra-group variance $\boldsymbol{\sigma_k^2} = \frac{1}{|G_k|}\sum_{c=1}^{|G_k|}(\mu_{c,k} - \mu_k)^2$, $\quad \boldsymbol{\mu_k} = \frac{1}{|G_k|}\sum_{c=1}^{|G_k|}\mu_{c,k}$
- A dynamic variance threshold $\boldsymbol{\sigma^2_{thresh}} = \boldsymbol{precentile}\left(\sigma_k^2, p\right)$, $\quad \boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$
- Groups for reassignment $\boldsymbol{G_{assign}} = \{G_k | \sigma_k^2 > \sigma^2_{thresh}, \quad \forall k \in \{1, 2 \ldots . K\}\}$
- Channel candidates for regrouping $\boldsymbol{F_{assign}} = \left\{\{f_{G_k} | \mathbb{I}(k \in G_{assign}) = 1\}\right\}$
- Group bounds – $\boldsymbol{L_k} = \left\lfloor (1-\alpha) \cdot C_{G_{def}} \right\rfloor$, $\boldsymbol{H_k} = \left\lfloor (1+\alpha) \cdot C_{G_{def}} \right\rfloor$, $\alpha \in (0, 1)$
- Final Group Normalization: $\boldsymbol{ConstrainKMeans}\left(F_{assign}, L_k, H_k\right) \cup G_{unassign}$

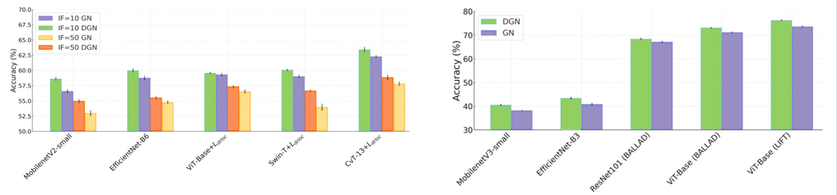


Experiments

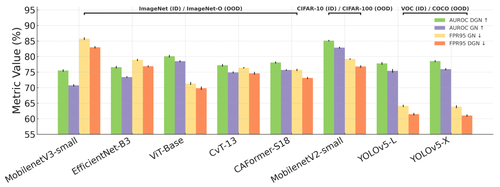
SOTA Performance



Preserving Minority-Class Features



Robustness to Out-of-Distribution (OOD) and corruptions



Data	Metric	Model	DGN	GN
CIFAR100-C	CER ↓	ResNext-29A [56]	**25.47**	26.33
		WRN40-2A [56]	**30.23**	31.87
ImageNet-C	CER ↓	ResNet-50 [56]	**41.52**	43.48
		ResNet-50A [56]	**32.07**	33.66
	mCE ↓	ViT-Base [15] (FAN-B-Hybrid + RSPC)	**45.74**	46.81
ImageNet-S	Mean IoU ↑	ViT-Base [11] (TEC_{MAE})*	**61.04**	59.79
		ViT-Base [29] (iBOT + SERE)*	**62.87**	61.23

Conclusions

DGN is the first normalization method to dynamically optimize group sizes and compositions through statistical awareness. By adapting to both spatial and temporal data variations, DGN preserves minority-class features, improves robustness to out-of-distribution inputs and corruptions, and delivers state-of-the-art performance.